



King's Research Portal

DOI:

[10.1027/1618-3169/a000309](https://doi.org/10.1027/1618-3169/a000309)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Vadillo Nistal, M. A., Blanco, F., Yarritu, I., & Matute, H. (2016). Single- and dual-process models of biased contingency detection. *Experimental Psychology*, 63(1), 3-19. <https://doi.org/10.1027/1618-3169/a000309>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Running Head: Biased contingency detection

IN PRESS: Experimental Psychology (Sept. 2015)

Single- and Dual-Process Models of Biased Contingency Detection

Miguel A. Vadillo^{1,2}, Fernando Blanco³, Ion Yarritu³, & Helena Matute³

¹Primary Care and Public Health Sciences, King's College London, UK

²Department of Experimental Psychology, University College London, UK

³Departamento de Fundamentos y Métodos de la Psicología, Universidad de Deusto, Bilbao, Spain

Mailing address:

Miguel A. Vadillo

Primary Care and Public Health Sciences

King's College London

Capital House, 42 Weston St., London SE1 3QD, UK

e-mail: miguel.vadillo@kcl.ac.uk

Abstract

Decades of research in causal and contingency learning show that people's estimations of the degree of contingency between two events are easily biased by the relative probabilities of those two events. If two events co-occur frequently, then people tend to overestimate the strength of the contingency between them. Traditionally, these biases have been explained in terms of relatively simple single-process models of learning and reasoning. However, more recently some authors have found that these biases do not appear in all dependent variables and have proposed dual-process models to explain these dissociations between variables. In the present paper we review the evidence for dissociations supporting dual-process models and we point out important shortcomings of this literature. Some dissociations seem to be difficult to replicate or poorly generalizable and others can be attributed to methodological artefacts. Overall, we conclude that support for dual-process models of biased contingency detection is scarce and inconclusive.

Keywords: associative models; cognitive biases; contingency learning; cue-density bias; dual-process models; illusory correlations; outcome-density bias; propositional models.

Contingency learning is the ability to detect that different events in the environment are statistically related. Classical and instrumental conditioning are probably the simplest and more popular examples of contingency learning. However, this ability is also an essential part of more sophisticated cognitive processes like language acquisition (Ellis, 2008), visual search (Chun & Turk-Browne, 2008), causal induction (Holyoak & Cheng, 2011), or categorization (Kruschke, 2008). Given the importance of these processes, it is hardly surprising that people tend to be very good at detecting statistical correlations and causal relations since their first years of life (Beckers, Vandorpe, Debeys, & De Houwer, 2009; Gopnik, Sobel, Schulz, & Glymour, 2001; Saffran, Aslin, & Newport, 1996).

Unfortunately, we are so eager to detect statistical patterns that we also tend to perceive them when they are absent (Chapman & Chapman 1969; Matute, 1996; Redelmeier & Tversky, 1996). Understanding how and why we misperceive contingency between unrelated events has become one of the most interesting topics of research in cognitive psychology (Gilovich, 1991; Vyse, 1997). It has been suggested that biased contingency detection might play a role in the development of pseudoscientific thinking, clinical errors, social stereotyping, and pathological behavior, among others (Hamilton & Gifford, 1976; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2014; Matute, Yarritu, & Vadillo, 2011; Orgaz, Estévez, & Matute, 2013; Reuven-Magril, Dar, & Liberman, 2008). From this point of view, basic research on the mechanisms underlying these biases can make a potential contribution to the improvement of debiasing and educational strategies aimed at counteracting them (Barberia, Blanco, Cubillas, & Matute, 2013; Matute et al., 2015).

In the present review, we focus on two specific biases that are assumed to distort our perception of contingency, namely, the cue-density bias and the outcome-density bias (Allan & Jenkins, 1983; López, Cobos, Caño, & Shanks, 1998; Wasserman, Kao, Van-

Hamme, Katagiri, & Young, 1996). In the following sections we briefly explain these two biases and their contribution to our understanding of why we perceive illusory correlations and why we infer causal relations between events that are actually independent. Although biases in contingency detection have been explored extensively for decades, there is still little consensus about their underlying mechanisms. Traditionally, they have been explained in terms of relatively simple single-process models that put the stress on basic learning and memory processes (Fiedler, 2000; López et al., 1998; Shanks, 1995). However, more recent theories have suggested that multiple processes are needed to fully understand biases in contingency detection (Allan, Siegel, & Hannah, 2007; Allan, Siegel, & Tangen, 2005; Perales, Catena, Shanks, & González, 2005; Ratliff & Nosek, 2010). In general, these theories fit very well with the increasing popularity of dual-process models in cognitive psychology (Sherman, Gawronski, & Trope, 2014).

The goal of the present paper is to assess critically the evidence for dual-process models of biased contingency learning. To summarize our review, we first present the basic methodology used to explore biases in contingency detection and the main (single- and dual-process) theories designed to explain them. Then, we present the results of a reanalysis of our own published work that suggests that the findings that support dual-process models were not replicated in our own data set, comprising data from 848 participants. In addition, computer simulations show that the use of insensitive dependent measures might explain some results that are typically interpreted in terms of dual-process models. Finally, we explore the experiments that have tested the predictions of dual-process models with implicit measures and we argue that the pattern of results is too heterogeneous to draw any firm conclusions. In light of this, we conclude that for the moment being it would be premature to abandon traditional, single-process models. Unless future research shows otherwise, these models still provide the best and simplest

framework to understand biases in contingency detection and to design successful debiasing strategies.

Biased contingency detection and illusory correlations

Imagine that you were asked to evaluate whether a new medicine produces an allergic reaction as a side effect. To accomplish this task, you are shown the individual medical records of a number of patients where you can find out whether each patient took the medicine and whether he/she suffered an allergic reaction. How should you assess the relation between taking the medicine and suffering the allergy? As shown in Figure 1A, to make this judgment you would need four pieces of information that can be summarized in a 2×2 contingency table. You would need to know how many patients took the medicine and suffered an allergy (cell *a*), how many patients took the medicine and did not suffer an allergy (cell *b*), how many patients did not take the medicine but suffered an allergy nevertheless (cell *c*), and, finally, how many patients did not take the medicine and did not suffer an allergy (cell *d*).

Based on this information, you could compute some measure of contingency and estimate whether or not that level of contingency is substantially different from zero. Although there are alternative ways to measure contingency, the Δp rule is usually considered a valid normative index (Allan, 1980; Cheng & Novick, 1992; Jenkins & Ward, 1965). According to this rule, if you want to assess the degree of contingency between a cue (e.g., taking a medicine) and an outcome (e.g., an allergy), then you need to compute

$$\Delta p = p(o|c) - p(o|\sim c), \quad (1)$$

where $p(o|c)$ is the probability of the outcome given the cue and $p(o|\sim c)$ is the probability of the outcome in the absence of the cue. As shown in Figure 1A, these two probabilities can be easily computed from the information contained in a contingency table. Positive values of Δp indicate that the probability of the outcome is higher when the cue is present

than when it is absent (see, e.g., Figure 1B). In contrast, negative values indicate that the probability of the outcome is reduced when the cue is present (e.g., Figure 1C). Finally, if the probability of the outcome is the same in the presence as in the absence of the cue, the value of Δp is always 0 (e.g., Figure 1D). In these latter cases, there is no contingency between cue and outcome.

When lay people are asked to estimate the contingency between two events, do their judgments agree with this normative rule? As we will see, the answer is both “yes” and “no”. To study how people detect contingency, researchers typically rely on a very simple task that has become a standard procedure in contingency learning research. During the task, participants are exposed to a series of trials in which a cue and an outcome may be either present or absent, and they are instructed to discover the relationship between both. As in our previous example, the cue can be a fictitious medicine taken by some patients and the outcome can be an allergic reaction. On each trial, participants are first shown information about whether a patient took the drug on a specific day and they are asked to predict whether or not they think that this patient will develop an allergic reaction. After entering a yes/no response, they receive feedback and they proceed to the next trial. Once they have seen the whole sequence of trials in random order, the participants are asked to rate their perceived strength of the relationship between the medicine and the allergic reaction.

The usual result is that participants’ judgments tend to covary with the objective drug-allergy contingency as measured by Δp (e.g., López et al., 1998; Shanks & Dickinson, 1987; Wasserman, 1990). Therefore, to some extent participants seem to be able to track the actual cue-outcome contingency. However, departures from the objective contingency are also observed. For instance, participants’ judgments tend to be biased by the marginal probability of the outcome, defined as the proportion of trials in which the

outcome is present, that is, $p(\text{outcome}) = (a + c) / (a + b + c + d)$. Figure 1E depicts an example where there is no contingency between cue and outcome, but the outcome tends to appear very frequently. In situations like this, participants tend to overestimate contingency (Allan & Jenkins, 1983; Allan et al., 2005; Buehner, Cheng, & Clifford, 2003; Lopez et al., 1998; Musca, Vadillo, Blanco, & Matute, 2010; Wasserman et al., 1996). Similarly, other things being equal, participants' judgments tend to covary with the marginal probability of the cue, defined as the proportion of trials in which the cue is present; that is $p(\text{cue}) = (a + b) / (a + b + c + d)$. Figure 1F represents an example where the probability of the cue is high, but there is no contingency between cue and outcome. Again, participants tend to overestimate contingency in situations like this (Allan & Jenkins, 1983; Matute et al., 2011; Perales et al., 2005; Vadillo, Musca, Blanco, & Matute, 2011; Wasserman et al., 1996). The biasing effects of the probability of the outcome and the probability of the cue are typically known as *outcome-* and *cue-density biases*. As the astute reader might guess, the most problematic situation is that in which both the probability of the outcome and the probability of the cue are large (Figure 1G). Participants seem to find particularly difficult to detect the lack of contingency in these cases (Blanco, Matute, & Vadillo, 2013).

It is interesting to note that biases akin to these have also been found in the social psychology literature on illusory correlations in stereotype formation (Hamilton & Gifford, 1976; Kurtzner & Fiedler, 2015; Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton, 2011). In these experiments, participants are shown information about the personality traits of members of two social groups. Across trials, participants see more information about one of the groups than about the other and they also see more people with positive traits than people with negative traits. Most importantly, the proportion of positive and negative traits is identical in both social groups. Therefore, there is no correlation between membership to the majority or the minority group and the quality

(positive vs. negative) of personality traits. As can be seen, if one assumes that social groups play the role of cues and that positive and negative traits play the role of outcomes, this situation is identical to the one represented in Figure 1G. Although there is no correlation between groups and traits, when participants are asked to rate the traits of both groups, they systematically tend to judge the majority group more favorably than the minority group. In other words, despite the absence of a real correlation, participants tend to associate the majority group with the most frequent (positive) personality traits and the minority group with the least frequent (negative) personality traits. The interesting point, for our current purposes, is that we can interpret illusory correlations as a combination of cue- and outcome-density biases, which means that this effect may be explained as a contingency learning phenomenon.

Single-process models of cue- and outcome-density biases

Traditionally, demonstrations of cue/outcome-density biases and illusory correlations have been explained in terms of simple associative processes analogous to those assumed to account for classical and instrumental conditioning (e.g., Alloy & Abramson, 1979; López et al., 1998; Matute, 1996; Murphy, Schmeier, Vallée-Tourangeau, Mondragón, & Hilton, 2011; Shanks, 1995; Sherman, Kruschke, Sherman, Percy, Petrocelli, & Conrey, 2009; Van Rooy, Van Overwalle, Vanhooymissen, Labiouse, & French, 2003). The associative learning rule proposed by Rescorla and Wagner (1972) provides the simplest example of this family of models. According to the Rescorla-Wagner model, when a cue is followed by an outcome, an association or link is formed between the representations of both stimuli. After each pairing, the strength of the association is assumed to increase or decrease according to the formula:

$$\Delta V_{C-O} = \alpha \cdot \beta \cdot (\lambda - V_{TOTAL}), \quad (2)$$

where ΔV_{C-O} is the increase in the strength of the cue-outcome association after that trial, α and β are learning rate parameters depending on the salience of the cue and the salience of the outcome, respectively, λ is a dummy variable coding whether the outcome was present or absent in that trial, and V_{TOTAL} is the sum of the associative strengths of all the potential cues of the outcome present on that trial. In addition to the target cue, a contextual cue is assumed to remain present in all trials. The association of the contextual cue with the outcome is also updated according to Equation 2.

To illustrate how this simple model accounts for cue- and outcome-density biases, in Figure 2 we show the predictions of the model when given as input the six contingencies depicted in Figures 1B-1G. The top panel shows the predictions of the model when the contingency is positive (1B), negative (1C) or zero (1D). As can be seen, eventually the strength of the cue-outcome association tends to converge to the true contingency, as defined by Δp . By the end of training, the model learns a positive association when the cue-outcome contingency is positive and a negative association when the cue-outcome contingency is negative. When the cue-outcome contingency is exactly zero, the associative strength of the cue also tends to move towards this value. Therefore, the model does a good job at explaining why people are good at detecting contingencies (see Chapman & Robbins, 1990; Danks, 2003; Wasserman, Elek, Chatlosh, & Baker, 1993). However, the model also predicts some systematic deviations from the true contingency. In the four conditions where the contingency is zero, depicted in the bottom panel, the model predicts an overestimation of contingency during the initial stages of learning. These overestimations are larger when the outcome (1E) or the cue (1F) are very frequent, and even larger when both of them are very frequent (1G). Therefore, the model also provides a nice explanation for cue- and outcome-density biases (Matute, Vadillo, Blanco, & Musca, 2007; Shanks, 1995; Vadillo & Luque, 2013).

Regardless of the merits and limits of associative models (Mitchell, De Houwer, & Lovibond, 2009; Shanks, 2010), for our present purposes, their most important feature is that, according to them, the same mechanism explains (a) why people are sensitive to contingency and also (b) why their judgments are also biased under some conditions. A single process accounts for accurate and biased contingency detection. As we will discuss below, this is the key feature of single-process models that distinguishes them from their dual-process counterparts.

It is interesting to note that this property is also shared by other early models of biased contingency detection that do not rely on associative learning algorithms. For example, instance-based models assume that each cue-outcome trial is stored in a separate memory trace in long-term memory (Fiedler, 1996, 2000; Meiser & Hewstone, 2006; Smith, 1991). Parts of these memory traces may be lost during the encoding process. In a situation like the one represented in Figure 1G, the loss of information has very little impact on the encoding of cell *a* events, because there are many redundant memory traces representing the same type of event. However, information loss can have a severe impact on the encoding of cells *c* and *d* events because there are fewer traces representing them. As a result, the information encoded in memory contains more (or better) information about frequent events (cells *a*) than about infrequent events (cells *c* and *d*). Therefore, information loss explains why participants tend to perceive a positive contingency whenever type *a* events are more frequent than other events in the contingency table. Most importantly, according to these models we do not need to invoke different mechanisms to explain the cases in which participants are sensitive to the actual contingency and the instances in which their judgments are biased. Accurate and biased contingency detection are supposed to arise from the same operating mechanisms. Therefore, from our point of view, instance-based theories also belong to the category of single-process models.

For our present purposes, propositional models can be considered yet another case of single-process models. In a thought-provoking series of papers, De Houwer and colleagues (De Houwer, 2009, 2014; Mitchell et al., 2009) have suggested that all instances of human contingency learning might depend exclusively on the formation and truth evaluation of propositions. In contrast to simple associations, propositions do not just represent that events in the environment are related to each other: They also qualify how they are related (Lagnado, Waldmann, Hagmayer, & Sloman, 2007). For instance, “cholesterol is a cause of cardiovascular disease” and “cholesterol is a predictor of cardiovascular disease” are different propositions. However, the difference between them cannot be represented in terms of a simple association. Although propositional models do not necessarily exclude the contribution of associative processes (see Moors, 2014), the representational power of propositions allows these models to explain aspects of learning that fall beyond the scope of simple associative models (De Houwer, Beckers, & Glautier, 2002; Gast & De Houwer, 2012; Zanon, De Houwer, & Gast, 2012). These ideas have not been formalized in a mathematical model, but nothing in their current formulation suggests that separate mechanisms would be needed to account for accurate and biased contingency detection. For our present purposes, the idea that all learning depends on the evaluation of propositions represents yet another example of a single-process model.

Dissociations and dual-process models

During the last decade, some researchers have abandoned these explanations in favor of more complex dual-process theories (Allan et al., 2005, 2007; Perales et al., 2005; Ratliff & Nosek, 2010). Although differing in the detail, the core idea of these proposals is that they call upon one mechanism to explain how people (correctly) track contingencies and a different mechanism to explain why their judgments are sometimes biased by cue

and outcome density. This proposal is based on the results of several experiments showing what appear to be systematic dissociations between different dependent measures.

Cue/outcome density biases and illusory correlations are typically assessed with a numerical causal or contingency rating that participants provide at the end of the experiment. As explained, these judgments show sensitivity to both actual contingency and to the biasing effects of cue and outcome density. However, according to these authors (Allan et al., 2005, 2007; Perales et al., 2005; Ratliff & Nosek, 2010), other dependent measures seem to be sensitive only to the actual contingency, showing no trace of cue- or outcome-density biases. These alternative measures are assumed to be relatively uninfluenced by higher-order reasoning processes, or at least, less influenced by them than the numerical judgments typically used as dependent variables. From this point of view, it follows naturally that there must be a very basic learning mechanism that explains how people accurately track contingencies and whose output can be directly observed in these dependent variables. In contrast, judgments are affected both by contingency and by cue/outcome density biases. Because measures that address directly the learning mechanism do not seem to be sensitive to biases, these must be operating through a different mechanism that influences judgments but not the original encoding of information. This is the reason why these models incorporate different processes to account for accurate and biased contingency detection. A schematic representation of the role of learning and judgment processes of biased contingency detection is offered in Figure 3.

Therefore, in dual-process models two different and successively operating mechanisms are invoked to explain why people are sensitive to contingency but they are also biased by the marginal probabilities of the cue and the outcome. The first mechanism would include basic encoding and retrieval processes that are highly sensitive to the

objective contingency. The information gathered by this mechanism would then feed forward to other mechanisms involved in judgment and decision-making processes. Biases would appear only at this latter stage.

These findings are certainly challenging for the theories of contingency detection discussed in the previous section, which in the absence of additional assumptions would typically anticipate similar effects in all dependent measures of contingency learning. Note, however, that dissociations are not a perfect basis to draw inferences about the presence of one or multiple systems. Borrowing an example from Chater (2003), following the logic of dissociations one might conclude that humans must have different digestive systems, because some people are allergic to prawns, while others are allergic to peanuts. Tiny differences in the way a single mechanism tackles similar problems might create the illusion that several mechanisms are involved in an operation that is actually best described in terms of a single-system process.

In spite of these concerns about the interpretation of dissociations, in the following sections we do not question this logic, but the reliability of the findings that support dual-process models of contingency learning. The evidence for these models stems from three papers published during the last decade (Allan et al., 2005; Perales et al., 2005; Ratliff & Nosek, 2010). Although their theoretical conclusions are quite consistent, the empirical findings reported in each of them are noticeably different. In the following sections we review each of them in turn and discuss their merits and shortcomings. To overview our criticisms, we argue that some of these findings seem to be poorly replicable or generalizable, while others are based on possibly faulty dependent measures.

Cue- and outcome-density biases in trial-by-trial predictions

The first piece of evidence suggesting that these biases are not observed in all dependent variables comes from an interesting experiment conducted by Allan et al.

(2005). Two groups of participants were instructed to discover the effect of a series of fictitious chemicals (playing the role of the cue) on the survival of a sample of bacteria in a petri dish (playing the role of the outcome). In each trial, participants saw whether or not the chemical was present in a sample and they were asked to predict whether or not the bacteria would survive. Immediately after entering their responses, they were informed of the outcome of the trial (i.e., whether the bacteria survived) and they proceeded to the next trial. At the end of training, participants were asked to rate to what extent the chemicals had a positive or a negative impact on the survival of bacteria, using a numerical scale from -100 (negative impact) to 100 (positive impact). The overall chemical-survival contingency was different for each group of participants. For one of them, the contingency was moderately positive ($\Delta p = .467$), while for the other one the contingency was always null ($\Delta p = .000$). Each participant was asked to complete three of these contingency-detection problems, all of them with the same overall contingency, but with different probabilities of the outcome. Therefore, the experiment relies on a 2×3 factorial design with contingency as a between-groups manipulation and probability of the outcome as a within-participant factor.

As the reader might expect, the first finding of Allan et al. (2005) was that the numerical ratings that participants provided at the end of each problem were influenced both by contingency and outcome-probability. That is to say, participants were able, in general, to track the objective degree of contingency between each of the chemicals and the survival of bacteria; however, their ratings were also biased by the probability of the outcome. This is a replication of the well-known outcome-density bias discussed in previous sections.

Most interestingly, Allan et al. (2005) found that other dependent measures seemed to be unaffected by outcome density, although they were sensitive to the overall cue-

outcome contingency. Specifically, Allan et al. used the discrete yes/no predictions made by participants in every trial to compute an alternative measure of their sensitivity to contingency. If the participant believes that there is a statistical connection between the chemicals and the survival of bacteria, then he or she should predict the survival (i.e., respond “yes” to the question of whether the bacteria would survive in the current trial) more frequently when the chemicals are present than when they are not. Following this reasoning, it would be possible to measure the extent to which a participant believes that there is a relationship between the cue and the outcome using the formula:

$$\Delta p_{\text{pred}} = p(\text{“yes”}|\text{cue}) - p(\text{“yes”}|\sim\text{cue}). \quad (3)$$

Note that this index is based on the same logic that underlies the computation of Δp in Equation 1, only that the real occurrence of the outcome is replaced by the outcome predictions made by the participant (see Collins & Shanks, 2002). Therefore, Δp_{pred} does not measure the objective contingency between cue and outcome, but it aims at measuring the subjective contingency that the participant perceives, as revealed by the trial-by-trial predictions made during training.

When using Δp_{pred} as their dependent variable, Allan et al. (2005) found that this measure was sensitive to cue-outcome contingency. However, it was absolutely unaffected by manipulations of the probability of the outcome. This result led them to conclude that there must be a stage of processing in which the contingency between cue and outcome or the target conditional probabilities have already been encoded but the outcome-density bias is still absent. Then, Δp_{pred} would provide an insight into this basic mechanism that encodes contingency in a format that is not yet influenced by the outcome-density bias. Given that the numerical judgments are influenced by both contingency and outcome-probability, this must mean that judgments are affected not only by the original encoding of information (sensitive to contingency but free from bias) but also by processes that take

place after encoding (see Figure 3). In other words, the outcome-density bias is not due to learning or encoding, but to more sophisticated processes related to judgment and decision making. Although Allan et al. (2005) explored only the outcome-density bias, they suggested that a similar approach might explain the complementary cue-density bias as well.

The results of Allan et al. (2005) and their interpretation are certainly appealing. If they proved to be reliable, they would pose insurmountable problems for single-process models that aim to explain cue- and outcome-density biases as learning effects. But, how strong is this evidence? To answer this question, we decided to reanalyze data from our own laboratory using the strategy followed by Allan et al. Specifically, we reanalyzed data from nine experimental conditions exploring the cue-density bias (originally published in Blanco et al., 2013, Matute et al., 2011, Vadillo et al., 2011; Yarritu & Matute, 2015; Yarritu, Matute, & Vadillo, 2014) and three experimental conditions exploring the outcome-density bias (originally published in Musca et al., 2010; Vadillo, Miller, & Matute, 2005). All these experiments were conducted using the standard experimental paradigm outlined above. In the original reports of those experiments we only analyzed the judgments that participants reported at the end of the training phase. However, we also collected trial-by-trial predictive responses to maintain participants' attention and to make sure that they were following the experiment. These responses can be used to compute the Δp_{pred} index using Equation 3. This allows us to compare the size of the bias observed in the Δp_{pred} scores with the size of the bias that we observed in judgments. On the basis of the results of Allan et al., one should expect a dissociation between these two measures. More specifically, cue- and outcome-density biases should have an effect on judgments, but not on Δp_{pred} .

In the following analyses, we included data from 848 participants tested in 12 conditions included in the articles mentioned in the previous paragraph.¹ Figure 4 plots the effect size (Cohen's d) of density biases on the Δp_{pred} index against the effect size of the same manipulation on judgments collected at the end of the experiment. As can be seen, overall, most experiments found clear evidence for density biases in Δp_{pred} . A random-effects meta-analysis yielded a statistically significant effect of size $d = 0.49$, 95% CI [0.08, 0.90], $z = 2.37$, $p = .018$. Therefore, overall our results do not replicate the original findings of Allan et al. (2005): Cue and outcome density do have an effect on Δp_{pred} scores. However, the confidence intervals plotted in Figure 4 show that the effect of density bias did not reach statistical significance in some experimental conditions. In two cases, the effect was even negative. A closer look at the data shows that in the rare occasions when biases were not observed in Δp_{pred} , they also tended to be absent or smaller than usual in judgments. A meta-regression confirmed that the effect size of biases on Δp_{pred} was moderated by the effect size of biases on judgments, $Q(1) = 5.54$, $p = .019$. These analyses suggest that cue- and outcome-density biases can be observed in Δp_{pred} and that, when they are absent, it is not due to a dissociation between judgments and Δp_{pred} , but to any other factor that affects both measures.

To show that these results are robust, we also analyzed the individual data of the participants tested in all the experimental conditions shown in Figure 4. Across all conditions, the cue-/outcome-density manipulation had an effect on Δp_{pred} , $t(846) = 6.20$, $p < .001$, $d = 0.43$. This result also held when data from experiments exploring cue- and outcome-density biases were analyzed separately: $t(659) = 4.97$, $p < .001$, $d = 0.37$, and $t(185) = 4.01$, $p < .001$, $d = 0.59$, respectively. Not surprisingly, the two dependent measures, judgments and Δp_{pred} , were significantly correlated, $r = .27$, $p < .001$, and this correlation remained significant when data from cue- and outcome-density biases were

analyzed independently: $r = .28, p < .001$, and $r = .24, p < .001$, respectively. Overall the data shown in Figure 4 and these additional analyses are inconsistent with the hypothesis that trial-by-trial predictions are unbiased by the probability of the cue/outcome or that radically different results are observed with judgments and trial-by-trial predictions. Thus, there is no need to postulate different mechanisms to account for judgments and for trial-by-trial predictions.

This being said, Figure 4 also reveals that some of our studies failed to find a significant effect of the cue/outcome density manipulation on Δp_{pred} . To some extent, this feature of our results can be considered a replication of Allan et al. (2005). But as seen in Figure 4 and in the previous analyses, this can hardly be considered strong evidence for a dissociation between judgments and Δp_{pred} . The fact that sometimes Δp_{pred} fails to yield significant results may be due to its reduced reliability compared to judgments. There are many situations where the use of unreliable or insensitive measures can produce patterns of results that look like dissociations but do not require a dual-process account (Shanks & St. John, 1994; Vadillo, Konstantinidis, & Shanks, in press). Consistent with this interpretation, there are good reasons why Δp_{pred} might be an imperfect index of contingency learning, as we discuss in subsequent sections.

Signal Detection Theory analyses of trial-by-trial predictions

Perales et al. (2005, Experiment 1) found a dissociation strikingly similar to the one reported by Allan et al. (2005). In the study of Perales et al., two groups of participants were exposed to several contingency-learning problems where they had to learn the relationship between the activation of a fictitious minefield and the explosion of enemy tanks. In each trial, participants were first presented with information about whether the minefield was active and were asked to predict by means of a yes/no response whether or not they thought that the tanks would explode in that trial. After entering their predictions,

they were given feedback and they proceeded to the next trial. For one of the groups, the contingency between the minefield and the explosions was always positive, $\Delta p = .50$, while the contingency was always zero for all the problems presented to the other group of participants. Within participants, the probability of the cue was manipulated with two levels, high (.75) and low (.25). At the end of each problem, participants were asked to rate the strength of the causal relation between the activation of the minefield and the explosion of tanks. Consistent with previous reports, Perales et al. found that these numerical judgments were sensitive not only to the contingency manipulation, but also to the cue-density manipulation. That is to say, for a specific level of contingency, judgments tended to vary with the probability of the cue, replicating the well-known cue-density bias. However, as in the case of Allan et al., Perales et al. found that other dependent measures, also computed from participants' trial by trial responses, were sensitive only to contingency and were largely immune to the cue-density bias.

Unlike Allan et al. (2005), Perales et al. (2005, Experiment 1) did not convert trial-by-trial predictions to Δp_{pred} but, instead, they computed two alternative measures inspired in Signal Detection Theory (SDT) analyses. One of them, the criterion for responding, was a measure of participants' overall tendency to predict that the outcome will occur. The second one, d' , was the discriminability index of SDT analyses and aimed at measuring participants' ability to discriminate when the outcome was more likely to appear and when it was less likely to appear. For reasons that will become obvious later, only this second measure is relevant for the present discussion. To compute the d' index, Perales et al. registered the 'hit rate' of each participant (i.e., the proportion of trials in which they correctly predicted that the outcome would occur, among all trials where the outcome was present) and their 'false alarm rate' (i.e., the proportion of trials in which they incorrectly

predicted the outcome, among all trials where the outcome was absent). Based on these two measures, d' can be easily computed as

$$d' = z(\text{hit rate}) - z(\text{false alarm rate}), \quad (4)$$

where z is the inverse of the normal cumulative density function. Crucially, this equation ignores completely whether participants made their predictions on cue-present or on cue-absent trials. The only important thing is whether they correctly predicted the outcome when it was going to happen and whether they incorrectly predicted the outcome when it was not going to happen. In other words, the d' index measures to what extent participants are good at discriminating when the outcome will be presented and when it will not. The rationale for using this index as a measure of participants' sensitivity to contingency is that, in principle, if participants have learned the correct cue-outcome contingency, they should be able to make more accurate predictions, and this should yield a higher d' .

The key finding of Perales et al. (2005, Experiment 1) was that participants' d' scores turned out to be sensitive just to the contingency manipulation, but not to the cue density manipulation. This parallels Allan et al.'s (2005) finding that Δp_{pred} was affected by manipulations of contingency, but not by manipulations of the probability of the outcome. Taken collectively, both experiments converged on the same idea: There are some dependent measures that reflect that participants have learned the cue-outcome contingency, but which nevertheless show no trace of cue- or outcome-density bias. This stands in stark contrast with the patterns of results found in numerical judgments, which are sensitive to both contingency and density biases.

Perales et al. discussed several dual-process accounts that could explain these dissociations. Although differing in the detail, all of them dovetailed with the idea that there is a basic encoding mechanism that tracks cue-outcome contingency in a format free from any cue- or outcome-density bias. The d' index would be a direct measure of this

unbiased learning process. The density biases observed in judgments must then be attributed to other mechanisms that intervene in later stages of processing. This account fits well with the general framework outlined in Figure 3.

However, a closer inspection of the results reported by Perales et al. (2005, Experiment 1) suggests that alternative interpretations are possible. A first striking feature of the results is that the manipulation of the probability of the cue did in fact seem to have an effect on d' , although this effect was only marginally significant ($p = .09$, $\eta_p^2 = .067$). The authors argued that this effect was “far too small to account for the significant effect of cause-density [i.e., cue-density] on judgments.” (p. 1109). However, this argumentation is only valid if one assumes that the validity, reliability, and sensitivity of d' as a dependent measure are comparable to those of judgments. If d' turned out to be less sensitive, then the smaller effect size of the cue-density effect found of d' would be very poor evidence for a dissociation.

Are there any reasons to suspect that d' is not a sensitive measure for cue- or outcome-density biases? We think so. The problems of d' as a measure of learning are particularly obvious in the two null-contingency conditions of Perales et al. (2005, Experiment 1). In those conditions, the probability of the outcome was always .50, no matter whether the cue was present or not. In this situation, there was nothing participants could do to predict the outcome successfully. The outcome and its absence were equally likely and the cue did not offer any information to make the outcome more predictable. Furthermore, this was the case in both the high cue-density and the low cue-density conditions: The probability of the cue was higher in one condition than in the other, but this did not change the fact that the outcome was equally unpredictable in both conditions. Therefore, it is not surprising that participants in both conditions got d' values indistinguishable from zero. Note that this does not mean that participants had the same

perception of contingency in those two conditions. It only means that in those particular conditions, nothing participants learn can help them make better outcome predictions as measured by d' . But crucially, if all participants produce a d' of zero regardless of their predictions, then their score cannot be used as a measure of their perception of contingency.

A computer simulation provides a simple means to illustrate the problems of d' as a measure of contingency learning. In the following simulation we computed d' for a large number of simulated participants exposed to the same contingencies used in Perales et al. (2005, Experiment 1). The labels PC-HD, PC-LD, NC-HD, and NC-LD shown in Figure 5 refer to the four experimental conditions tested by Perales et al. These labels denote whether contingency was positive (PC) or null (NC) and whether the density of the cue was high (HD) or low (LD). For each experimental condition, we computed how many correct and incorrect responses a simulated participant would get. Then, we used this information to compute the d' of that simulated participant using Equation 4.²

For each condition, we simulated 1,000 participants with nine different response distributions. The rationale for simulating a wide variety of response strategies is that, if d' is a valid measure of learning, this index should adopt different values when participants behave differently. Imagine that one participant learns that there is positive contingency between the cue and the outcome. This participant should say “yes” very frequently when asked whether the outcome will follow the cue and he/she should say “no” very frequently when asked whether the outcome will appear in a trial in which the cue was absent. Now imagine a second participant who learns that there is no contingency between cue and outcome. This participant will be just as likely to predict the outcome in cue-present and in cue-absent trials. If d' is a good measure of contingency learning, these two participants should get different d' scores. In contrast, if participants who act on the basis of different

beliefs about the cue-outcome contingency obtain the same d' score, this would imply that d' is not a valid measure of contingency detection. To minimize the impact of sampling error on the exact number of hits and false alarms, we run 1,000 simulations for each combination of experimental condition and response distribution.

Each of the nine series of data shown in Figure 5 refers to a different response distribution (e.g., 25/75, 50/50...). The first number (25, 50 or 75) refers to the probability of predicting the outcome in the presence of the cue and the second number (also 25, 50 or 75) refers to the probability of predicting the outcome in the absence of the cue. For instance, a simulated participant with response strategy 75/25 would predict the outcome with probability .75 if the cue was present and with probability .25 if the cue was absent (which is consistent with the belief in a positive contingency). Similarly, a simulated participant with response strategy 75/75 would predict the outcome with probability .75 regardless of whether the cue is present or absent (which is consistent with the belief in a null contingency).

Consistent with our previous discussion, Figure 5 shows that all simulated participants got almost identical d' scores in the null contingency conditions (right-most half of the figure). In other words, the d' scores obtained by those participants reveal absolutely nothing about their pattern of performance. In these null contingency conditions, a participant who acted as if there were a positive cue-outcome contingency (e.g., 75/25) would receive virtually the same score as a participant who acted as if contingency were negative (e.g., 25/75). Similarly, a participant who were very prone to predicting the outcome (e.g., 75/75) and a participant who were very reluctant to predict the outcome (e.g., 25/25) would obtain similar d' scores. These predictions do not differ across conditions with different cue-densities. This confirms our suspicion that the d' scores obtained by Perales et al. (2005, Experiment 1) in the NC conditions tell us nothing

about participants performance, let alone about their beliefs regarding the cue-outcome contingency or their outcome expectancies.

In contrast, the left-hand side of Figure 5 shows that d' can be a sensitive measure of the perception of contingency in the positive contingency (PC) conditions tested by Perales et al. In these conditions, different patterns of responding do give rise to different d' scores, suggesting that d' can reveal something about participants' beliefs or about their response strategies. Therefore, it is only in these conditions with positive contingencies that one can expect to measure differences in performance with d' . It is interesting to note that visual inspection of the data reported by Perales et al. (2005) confirms that the trend towards a cue-density bias was stronger in the PC condition than in the NC condition. When one takes into account that one half of the experiment (the two NC conditions) is affected by a methodological artifact, it becomes less surprising that the cue-density manipulation only had a marginally significant effect on d' .

Different processes or different strategies?

This being said, we note that in the PC conditions the cue-density effect observed in d' still looks relatively small, compared to the large effect found in participants' judgments. Going back to the reanalysis of our own studies presented on our previous section, we also found there that, in some occasions, cue/outcome density manipulations seemed to have a stronger effect on judgments than on Δp_{pred} . Based on this evidence, it appears that, in general, all the dependent measures computed from trial-by-trial predictions (either Δp_{pred} or d') are less sensitive than numerical judgments. Is there any reason why these dependent measures should be less reliable? As we will show below, we suspect that the data reported by Allan et al. (2005) and Perales et al. (2015) provide an interesting insight into this question.

Imagine that two participants, A and B, have been exposed to exactly the same sequence of trials and that, as a result, they end up having the same beliefs about the relationship between a cue and an outcome. For instance, imagine that both of them have learned that the probability of the outcome given the cue is .75 and that the probability of the outcome given the absence of the cue is .25. This means that both of them would believe, implicit or explicitly, that there is a moderate positive contingency between cue and outcome (i.e., that the outcome is more likely to appear in the presence of the cue than in its absence). Now, let's assume that both participants are presented again with a series of trials where the cue is present or absent and they are asked to predict whether or not the outcome will be presented in each trial. Participant A might consider that, because the probability of the outcome given the cue is .75, he should predict the outcome in roughly 75% of the trials where the cue is present. And, similarly, because the probability of the outcome given the absence of the cue is .25, he predicts the outcome in approximately 25% of the trials where the cue is absent. The behavior of this participant would show what researchers call 'probability matching' (Nies, 1962; Shanks, Tunney, & McCarthy, 2002; Tversky & Edwards, 1966), that is, his predictions would match the probabilities seen (or perceived) in the environment.

Now imagine that Participant B is asked to do the same task. But Participant B has a different goal in mind: He wants to be right as many times as possible. If the outcome appears 75% of the times when the cue is present, then predicting the outcome on 75% of the trials is not a perfect strategy. If he did that, on average, he would be right on 56.25% of the trials (i.e., $.75 \times .75$). In contrast, if he always predicts the outcome when the cue is present, he will be correct 75% of the times (i.e., $1.00 \times .75$). If he wants to maximize the number of correct outcome predictions, this is a much more rational strategy. Following the same logic, if the probability of the outcome in the absence of the cue is .25, it makes

sense to always predict the absence of the outcome. Doing that will allow him to be correct on 75% of the trials. Our point is that if a participant wants to maximize the number of correct predictions, he will predict the outcome whenever he thinks that its probability is higher than .50 and he will predict the absence of the outcome in any other case. Not surprisingly, research with human and nonhuman animals shows that maximization is a typical response strategy in many situations (e.g., Unturbe & Coromias, 2007).

Interestingly, although both participants, A and B, base their responses on the same ‘beliefs’, their behavior is radically different because they pursue different goals. This has important implications for our review of the results reported by Allan et al. (2005) and Perales et al. (2005). If trial-by-trial predictions do not only depend on the perceived contingency but also on response strategies like probability matching or maximization, then any dependent variable computed from them (like Δp_{pred} or d') will be necessarily noisy and unreliable. This is particularly problematic when many participants rely on a maximization strategy. For instance, if we computed Δp_{pred} for Participant A in our previous example, that would yield an approximate value of .50, which reflects quite well his beliefs about the contingency between cue and outcome. However, if we computed Δp_{pred} for Participant B, this would yield a value of 1.00, which is a gross overestimation of his true beliefs. These two participants would also receive different d' scores.

In the case of Perales et al. (2005, Experiment 1), there is clear evidence of probability maximization. The data reported in the Appendix of Perales et al. confirm that participants in the PC group predicted the outcome in around 92-93% of cue-present trials and in 12-13% of cue absent trials. In the case of Allan et al. (2005) there is no obvious evidence for maximization in their non-contingent condition, but there appears to be such a trend in the contingent condition. Their Figure 5 suggests that although the actual probability of the outcome given the cue varied from .567 to .900, participants predicted

the outcome on 70-90% of cue-present trials. Similarly, although the probability of the outcome in the absence of the cue ranged from .100 to .667, participants predicted the outcome in 10-30% of cue-absent trials. This pattern is perhaps less extreme than the one found in Perales et al. (2005), but it does nevertheless suggest that many of their participants must have used a maximization strategy. In either case, this strategy makes Δp_{pred} and d' less sensitive to any manipulation. The lack of sensitivity might explain why they failed to find any effect of cue and outcome density on these dependent variables.

It is interesting to note that participants are more likely to become ‘maximizers’ in relatively long experiments, which provide more opportunities to develop optimal response strategies (Shanks et al., 2002). This might explain the diverging results obtained by Allan et al. (2005) and Perales et al. (2005) and our own experiments. In their experiments, cue and outcome density were manipulated within-participants. To accomplish this, all participants had to complete the contingency learning several times. In contrast, in our experiments all the manipulations were conducted between-groups, reducing substantially the length of the experiment and, consequently, the opportunities to develop sophisticated response strategies like maximization.

Illusory correlations in the Implicit Association Test

Dissociations between judgments and trial-by-trial predictions are not the only piece of evidence in favor of dual-process models of biased contingency detection. This approach received convergent support from a recent study by Ratliff and Nosek (2010) that found a similar dissociation between different measures of illusory correlations in stereotype formation, suggesting that two or more processes might also be involved in this effect.

As explained in previous sections, most experiments on illusory correlations in stereotype formation rely on a fairly standard procedure (Hamilton & Gifford, 1976).

Participants are presented with positive and negative traits of members of two different social groups on a trial-by-trial basis. Crucially, there are more members of one group (majority) than of the other (minority) and, regardless of group, most of members show positive traits. Although the proportion of positive and negative traits is identical for the majority and the minority groups, people tend to make more positive evaluations of the majority group when asked to judge both groups at the end of training. Illusory correlations in stereotypes and cue/outcome density biases have been explored in quite different literatures, but both effects are clearly related and can be explained by the same or very similar models (Murphy et al., 2011; Sherman et al., 2009; Van Rooy et al., 2003).

Illusory correlations are typically assessed by means of numerical ratings (similar to judgments in contingency detection experiments) or by asking participants to recall which positive or negative traits were observed in the majority or the minority group. However, Ratcliff and Nosek (2010) wondered whether the same illusory correlation effects would be found in an alternative test that is supposed to provide a cleaner measure of the underlying attitudes of participants: The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). Unlike traditional questionnaires, the IAT is a reaction-time test that is traditionally assumed to measure implicit attitudes with little interference from higher order cognitive processes (De Houwer, Teige-Mocigmba, Spruyt, & Moors, 2009; Gawronski, LeBel, & Peters, 2007; Nosek, Hawkins, & Frazier, 2011). Following this idea, if illusory correlations require the operation of reasoning or inferential processes, then they should not be observed in the IAT. In contrast, if only very elemental associative processes are responsible for illusory correlations, then the IAT should be able to detect them.

Ratcliff and Nosek (2010) found the standard illusory correlation effect in the responses to the explicit questionnaire. However, there was no hint of the effect in the IAT scores in any of their two experiments. Most importantly, the absence of effects cannot be

attributed to the lack of validity of the IAT: The IAT was sensitive to the valence of the majority and the minority groups when there was a correlation between membership to one of them and the positive or negative personality traits. Nor can they be attributed to a lack of statistical power, given that the results were replicated in an online experiment with almost 900 participants.

Ratliff and Nosek (2010) interpreted these results in terms of a dual-process model (Gawronski & Bodenhausen, 2006) surprisingly similar to that invoked by Allan et al. (2005) and Perales et al. (2005) in the domain of contingency learning. According to them, the IAT would be only sensitive to the original encoding of associations in memory. From this point of view, the fact that performance in the IAT was unaffected by the illusory-correlation manipulation indicates that participants in their experiment correctly learned that there was no correlation between belonging to one group or the other and having positive or negative personality traits. Therefore, the illusory correlation effect observed in explicit judgments must have been due to additional higher-order cognitive processes that took place on a later stage, and not to the original learning mechanism responsible for the initial encoding of the information (see Figure 3). As in the case of the results reported by Allan et al. and Perales et al., these results pose problems for any single-process model that assumes that illusory correlations are the product of the same mechanisms responsible for the detection of contingency. If a single process were responsible for both sensitivity to contingency and for density biases, why should an IAT be sensitive to one of these manipulations (contingency) but not to the other (density)?

Before drawing any conclusion, it is convenient to review all the available evidence regarding this dissociation. Until recently, the study conducted by Ratliff and Nosek (2010) was the only experimental work that had tried to detect illusory correlations with the IAT. However, the latest attempt to replicate this result using a similar methodology has failed

to find any dissociation between explicit measures and the IAT. Using a very similar procedure to Ratliff and Nosek, Carraro, Negri, Castelli, and Pastore (2014) did find an illusory correlation on the IAT, showing that the original dissociation was either not reliable or, more likely, not generalizable to similar but not identical conditions. In a similar vein, a recent experiment conducted in our laboratory (Vadillo, De Houwer, De Schryver, Ortega-Castro, & Matute, 2013) found an outcome-density effect using the IAT. The divergences with Ratliff and Nosek are less surprising in this case, because Vadillo et al. used a radically different design and procedure. However, this discrepancy converges to the idea that the failure of the IAT to detect illusory correlations might not be a generalizable result.

To better illustrate the results found with the IAT, Figure 6 depicts a forest plot with the divergent results of these studies.³ As can be seen, the only firm conclusion that can be drawn on the basis of this evidence is that the results are strikingly variable. In fact, the meta-analysis of these studies yielded an unusually large heterogeneity, $Q(4) = 134.18, p < .001$. Even the replication that Carraro et al. (2014) reported in their general discussion yielded results notably distant from those of their main study, although both of them were statistically significant. This variability suggests that whether or not illusory correlations are observed in the IAT probably depends on a number of moderators that we still ignore. As shown in Figure 6, the 95% confidence interval of the random effects model includes zero. An advocate of dual-process models might claim that this null-result of the meta-analysis supports the claim that illusory correlations are not found in implicit measures like the IAT. However, the confidence interval does not just include zero: It extends over a large number of positive effect sizes. On the basis of the collective evidence, any value from -0.13 to a massive 1.16 could be an accurate estimate of the average Cohen's d . We doubt that this evidence is clear or robust enough to abandon single-process models of

illusory correlations, which offer a simple and parsimonious explanation for a large body of data (Allan, 1993; Fiedler, 2000; López et al., 1998; Shanks, 1995). Even more so, if we keep in mind that the only converging evidence from Allan et al. (2005) and Perales et al. (2005) is open to criticism.

General Discussion

In the previous sections we have reviewed the studies that have found dissociations in cue- and outcome-density effects across dependent variables (Allan et al., 2005; Perales et al., 2005; Ratliff & Nosek, 2010). A common result of these experiments is that there are some dependent measures that only show sensitivity to contingency (e.g., Δp_{pred} , d' , or IAT scores), while other dependent measures (e.g., contingency judgments) show sensitivity to both contingency and cue/outcome density. On the basis of this evidence, it has been suggested that two separate mechanisms are needed to explain a) why people are able to learn the correct cue-outcome contingencies and b) why their judgments are influenced not only by contingency but also by the overall probabilities of the cue and the outcome. However, on closer inspection, it appears that this evidence might not be reliable enough to justify this theoretical interpretation. A review of the available evidence from our laboratory and also from other research groups shows that some of these results do not seem to be replicable or do not generalize easily to similar experimental settings. In our experiments, cue and outcome density manipulations seem to have a significant impact on all dependent measures. Similarly, our simulations show that some of the dissociations previously reported in the literature might be due to simple methodological artefacts resulting from aspects of the design and the computation of the dependent variable.

Given the lack of strong evidence in favor of these dual-process accounts, we think that it is premature to abandon the idea that sensitivity to contingency and density biases are both attributable to the operation of a single mechanism. As mentioned in the

introduction, this idea is an essential feature of many associative models that were originally invoked to account for cue- and outcome-density effects (López et al., 1998; Shanks, 1995; Sherman et al., 2009). It is also a central feature of alternative models of biased contingency detection, such as instance-based models (Fiedler, 1996; Meiser & Hewstone, 2006; Smith, 1991). Beyond the specific details of these models, their common feature is that they all share the assumption that the same mechanism that is responsible for detecting and encoding the relationship between cues and outcomes is also responsible for density biases. In other words, there is no level of representation in the cognitive system where contingency information is represented in a format that is free from cue- or outcome-density biases.

Note that, although we favor single-process models of biased contingency detection, we do not ignore the fact that different processes might contribute to each of the dependent variables used in this kind of research. We do not question the idea that there are manipulations that might affect one dependent variable without affecting others. In fact, part of our own research has been directed at showing that judgments of contingency can vary considerably depending on seemingly minor procedural details like the wording of the test question (Matute, Vegas, & De Marez, 2002; Vadillo & Matute, 2007; Vadillo et al., 2005; see also Crocker, 1982; De Houwer et al., 2007; Perales & Shanks, 2008). Moreover, part of our discussion of Perales et al. (2005) relies on the idea that participants sometimes adopt response strategies that might mask their true perception of contingency. What we are suggesting here is that there are no strong reasons to assume that there are two different levels of representation of contingency information, one that closely mirrors objective contingencies and one where that information is biased by factors like the probability of the cue or the probability of the outcome. In the absence of stronger evidence, it appears more parsimonious to assume that a single representation is learned during contingency

learning experiments and that this representation is biased by cue and outcome density at the encoding level.

Dual-process models of biased contingency detection share some ideas with other dichotomist models of cognition (Evans & Over, 1996; Kahneman, 2011; Osman, 2004; Sloman, 1996; Stanovich & West, 2000). There is a crucial difference, though, between the dual models reviewed in this article and other dual-process models. Traditionally, dual models have tended to explain biases and cognitive illusions attributing them to very simple cognitive mechanisms that operate effortlessly and in a relatively automatic manner, usually related to simple encoding or retrieval processes. More complex cognitive mechanisms, usually strategic processes related to judgment and decision making, have usually been invoked to explain why people are sometimes able to overcome the harmful effect of these biases and intuitive reactions (e.g., Kahneman, 2011). Interestingly, the dual process models of biased contingency detection that we review here make the opposite interpretation: Correct contingency detection is attributed to the operation of basic encoding and retrieval processes, while biased judgments are attributed to more sophisticated judgment and decision-making processes. We might even say that these models present a benign view of biases in contingency detection: Although people might show a bias in their judgments, deep inside their cognitive system there is some level of representation where information is represented accurately (a similar perspective can be found in De Neys, 2012).

The question of whether biases in contingency detection are due to basic encoding and retrieval processes or whether they reflect the operation of judgment and decision-making processes is not only important from a theoretical point of view. As mentioned above, it has been suggested that these biases might contribute to the development of superstitious and pseudoscientific thinking (Gilovich, 1991; Lindeman & Svedholm, 2012;

Matute et al., 2011, 2015; Redelmeier & Tversky, 1996; Vyse, 1997). Given the societal costs of these and other biases, cognitive psychologists have started to develop a number of interventions and guidelines for protecting people from cognitive biases (Barberia et al., 2013; Lewandowsky, Ecker, Seifer, Schwarz, & Cook, 2012; Lilienfeld, Ammirati, & Landfield, 2009; Schmaltz & Lilienfeld, 2014). Interventions designed to reduce biases can only be successful to the extent that they are based on an accurate view of their underlying mechanisms. If the underlying information was somehow encoded in a bias-free format, as suggested by dual-process models of contingency learning, these beliefs should be relatively easy to modify. Teaching people how to use more rationally the information and intuitions they already have should suffice to overcome these biases. This prediction stands in stark contrast with the well-known fact that superstitions are difficult to modify or eradicate (Arkes, 1991; Lilienfeld et al., 2009; Nyhan, Reifler, Richey, & Freed, 2014; Pronin, Gilovich, & Ross, 2004; Smith & Slack, 2015). We think that the persisting effect of biases is more consistent with the view that these beliefs are hardwired in the way people encode information about the relationship between events, as suggested by single-process models. Based on the evidence we have discussed so far, it seems safe to suggest that attempts to debias superstitions and misperceptions of contingency should include teaching people to look for unbiased information.

References

- Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgement tasks. *Bulletin of the Psychonomic Society*, 15, 147-149.
- Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, 114, 435-448.
- Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, 14, 381-405.
- Allan, L. G., Siegel, S., & Hannah, S. (2007). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology*, 60, 482-495.
- Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior*, 33, 250-263.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgements of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, 108, 441-485.
- Alonso, E., Mondragón, E., & Fernández, A. (2012). A Java simulator of Rescorla and Wagner's prediction error model and configural cue extensions. *Computer Methods and Programs in Biomedicine*, 108, 346-355.
- Arkes, H. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110, 486-498.
- Barbería, I., Blanco, F., Cubillas, C. P., & Matute, H. (2013). Implementation and assessment of an intervention to debias adolescents against causal illusions. *PLoS ONE*, 8, e71303.
- Beckers, T., Vandenorpe, S., Debeys, I., & De Houwer, J. (2009). Three year-olds' retrospective revaluation in the blicket detector task: Backward blocking or recovery from overshadowing? *Experimental Psychology*, 56, 27-32.

- Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learning & Behavior*, 41, 333-340.
- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1119-1140.
- Carraro, L., Negri, P., Castelli, L., & Pastore, M. (2014). Implicit and explicit illusory correlation as a function of political ideology. *PLoS ONE*, 9, e96312.
- Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, 18, 537-545.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Chater, N. (2003). How much can we learn from double dissociations? *Cortex*, 39, 167-169.
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, 99, 365-382.
- Chun, M. M., & Turke-Browne, N. B. (2008). Associative learning mechanisms in vision. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 209-245). New York: Oxford University Press.
- Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition*, 30, 1138-1147.
- Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, 8, 214-220.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47, 109-121.

- De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, 37, 1-20.
- De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, B., & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 530-541). New York, NY: Guilford Press.
- De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, 55A, 965-985.
- De Houwer, J., Teige-Mocigmba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135, 347-368.
- De Houwer, J., Vandrope, S., & Beckers, T. (2007). Statistical contingency has a different impact on preparation judgements than on causal judgements. *Quarterly Journal of Experimental Psychology*, 60, 418-432.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7, 28-38.
- Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In P. Robinson and N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, U.K.: Psychology Press.
- Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, 103, 193-214.

- Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology*, 4, 25-58.
- Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *Quarterly Journal of Experimental Psychology*, 65, 1657-1674.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692-731.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181-193.
- Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York, NY: Free Press.
- Gopnik, A., Sobel, D., Schulz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37, 620-629.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, 12, 392-407.
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135-163.

- Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, 79, 1-17.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss, Giroux.
- Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 267-301). New York: Cambridge University Press.
- Kutzner, F. L., & Fiedler, K. (2015). No correlation, no evidence for attention shift in category learning: Different mechanisms behind illusory correlations and the inverse base-rate effect. *Journal of Experimental Psychology: General*, 144, 58-75.
- Lagnado, D.A., Waldmann, M. R., Hagmayer, Y., & Sloman, S.A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154-172). Oxford: Oxford University Press.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- Lewandowsky, S., Ecker, U. K. H., Seifer, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13, 106-131.
- Lilienfeld, S. C., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, 4, 390-398.
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, 9, 355-387.

- Lindeman, M., & Svedholm, A. (2012). What's in a term? Paranormal, superstitious, magical and supernatural beliefs by any other name would mean the same. *Review of General Psychology*, 16, 241-255.
- López, F. J., Cobos, P. L., Caño, A., & Shanks, D. R. (1998). The rational analysis of human causal and probability judgment. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 314-352). Oxford: Oxford University Press.
- Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychological Science*, 7, 289-293.
- Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, 6, 888.
- Matute, H., Vadillo, M. A., Blanco, F., & Musca, S. C. (2007). Either greedy or well informed: The reward maximization – unbiased evaluation trade-off. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the European Cognitive Science Conference* (pp.341-346). Hove, UK: Erlbaum.
- Matute, H., Vegas, S., & De Marez, P-J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 714-725.
- Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, 102, 392-405.
- Meiser, T., & Hewstone, M. (2006). Illusory and spurious correlations: Distinct phenomena or joint outcomes of exemplar-based category learning? *European Journal of Social Psychology*, 36, 315-336.
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32, 183-246.

- Moors, A. (2014). Examining the mapping problem in dual process models. In J. W. Sherman, B. Gawronski, B., & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 20-34). New York, NY: Guilford Press.
- Murphy, R. A., Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *Quarterly Journal of Experimental Psychology*, 64, 24-40.
- Musca, S. C., Vadillo, M. A., Blanco, F., & Matute, H. (2010). The role of cue information in the outcome-density effect: evidence from neural network simulations and a causal learning experiment. *Connection Science*, 22, 177-192.
- Nies, R. C. (1962). Effects of probable outcome information on two choice learning. *Journal of Experimental Psychology*, 64, 430-433.
- Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, 15, 152-159.
- Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, 133, 1-8.
- Orgaz, C., Estevez, A. & Matute, H. (2013). Pathological gamblers are more vulnerable to the illusion of control in a standard associative learning task. *Frontiers in Psychology*, 4, 306.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11, 988-1010.
- Perales, J. C., Catena, A., Shanks, D. R., & González, J. A. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: A signal detection theory approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1105-1120.

- Perales, J. C., & Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1482-1494.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111, 781-799.
- Ratliff, K. A., & Nosek, B. A. (2010). Creating distinct implicit and explicit attitudes with an illusory correlation paradigm. *Journal of Experimental Social Psychology*, 46, 721-728.
- Redelmeier, D. A., & Tversky, A. (1996). On the belief that arthritis pain is related to the weather. *Proceedings of the National Academy of Sciences*, 93, 2895-2896.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- Reuven-Magril, O., Dar, R., & Liberman, N. (2008). Illusion of control and behavioral control attempts in obsessive-compulsive disorder. *Journal of Abnormal Psychology*, 117, 334-341.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Schmaltz, R., & Lilienfeld, S. O. (2014). Hauntings, homeopathy, and the Hopkinstown goblins: Using pseudoscience to teach scientific thinking. *Frontiers in Psychology*, 5, 336.
- Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology*, 48A, 257-279.

- Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, 61, 273-301.
- Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 21: Advances in research and theory* (pp. 229-261). San Diego, CA: Academic Press.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367-447.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15, 233-250.
- Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.) (2014). *Dual-process theories of the social mind*. New York: Guilford Press.
- Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology*, 96, 305-323.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Smith, B. W., & Slack, M. B. (2015). The effect of cognitive debiasing training among family medicine residents. *Diagnosis*, 2, 117-121.
- Smith, E. R. (1991) Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, 27, 107-123.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.

- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral & Brain Sciences*, 23, 645-665.
- Tangen, J. M., & Allan, L. G. (2004). Cue interaction and judgments of causality: Contributions of causal and associative processes. *Memory & Cognition*, 32, 107-124.
- Tversky, A., & Edwards, W. (1966). Information versus reward in binary choices. *Journal of Experimental Psychology*, 71, 680-683.
- Unturbe, J., & Corominas, J. (2007). Probability matching involves rule-generating ability: A neuropsychological mechanism dealing with probabilities. *Neuropsychology*, 21, 621-630.
- Vadillo, M. A., De Houwer, J., De Schryver, M., Ortega-Castro, N., & Matute, H. (2013). Evidence for an illusion of causality when using the Implicit Association Test to measure learning. *Learning and Motivation*, 44, 303-311.
- Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (in press). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-015-0892-6
- Vadillo, M. A., & Luque, D. (2013). Dissociations among judgments do not reflect cognitive priority: An associative explanation of memory for frequency information in contingency learning. *Canadian Journal of Experimental Psychology*, 67, 60-71.
- Vadillo, M. A., & Matute, H. (2007). Predictions and causal estimations are not supported by the same associative structure. *Quarterly Journal of Experimental Psychology*, 60, 433-447.
- Vadillo, M. A., Miller, R. R., & Matute, H. M. (2005). Causal and predictive-value judgments, but not predictions, are based on cue–outcome contingency. *Learning & Behavior*, 33, 172-183.

- Vadillo, M. A., Musca, S. C., Blanco, F., & Matute, H. (2011). Contrasting cue-density effects in causal and prediction judgments. *Psychonomic Bulletin & Review*, 18, 110-115.
- Van Rooy, D., Van Overwalle, F., Vanhooymissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review*, 110, 536-563.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistics Software*, 36, 1-48.
- Vyse, S. (1997). *Believing in magic: The psychology of superstition*. New York: Oxford University Press.
- Wasserman, E. A. (1990). Detecting response-outcome relations: Toward an understanding of the causal texture of the environment. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 26 (pp. 27-82). San Diego, CA: Academic Press.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: The role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 174-188.
- Wasserman, E. A., Kao, S.-F., Van-Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation*, Vol. 34: *Causal learning* (pp. 207-264). San Diego, CA: Academic Press.
- Yarritu, I., & Matute, H. (2015). Previous knowledge can induce an illusion of causality through actively biasing behavior. *Frontiers in Psychology* 6, 389.
- Yarritu, I., Matute, H., & Vadillo, M. A. (2014). Illusion of control: The role of personal involvement. *Experimental Psychology*, 61, 38-47.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, 43, 155-165.

Author Note

The authors were supported by Grant PSI2011-26965 from Dirección General de Investigación of the Spanish Government and Grant IT363-10 from Departamento de Educación, Universidades e Investigación of the Basque Government. We are indebted to José Perales and David Shanks for their valuable comments on earlier versions of this article. Correspondence concerning this article should be addressed to Miguel A. Vadillo, King's College London, Department of Primary Care and Public Health Sciences, Capital House, 42 Weston St., London SE1 3QD, United Kingdom. E-mail: miguel.vadillo@kcl.ac.uk

Footnotes

¹ Blanco et al. (2013) reported two experiments, each of them including two conditions where the effect of the cue-density manipulation was tested. Cue-density was also manipulated in Matute et al. (2011), Vadillo et al. (2011), and Yarritu and Matute (2015, Experiment 2). The latter contributed to our analyses with two experimental conditions. Yarritu et al. (2014) reported two experiments manipulating cue density, but due to their design requirements, trial-by-trial predictions were only requested in the yoked condition of Experiment 1. In that condition the probability of the cue could actually adopt any value from 0 to 1. Following the original data analysis strategy of Yarritu et al. (2014), we categorized participants in the “low probability of the cue” condition if they belonged to the one third of the sample with the lowest probability of the cue, and we categorized them in the “high probability of the cue” condition if they belonged to the one third of the sample with the highest probability of the cue. As mentioned above, we also included in our analyses three experimental conditions exploring the outcome-density bias. One of them originally reported by Vadillo et al. (2005, Experiment 3, Group 0.50-0.00 vs. Group 1.00-0.50) and two reported by Musca et al. (2010).

In most of these experiments participants were asked to provide only one judgment at the end of training. However, in Matute et al. (2011) and Vadillo et al. (2005, 2011) they were asked to provide several judgments. In the case of Matute et al. (2011) and Vadillo et al. (2011), we included in the analyses the judgment that yielded the stronger cue-probability bias in each experiment. If anything, selecting judgments that show strong biases should make it easier to observe any potential dissociation between judgments and trial-by-trial predictions. In the case of Vadillo et al. (2005) the largest effect of outcome-density was observed for prediction judgments, but this cannot be considered a bias

(because it is normatively appropriate to expect the outcome to happen when its probability is very large; see De Houwer, Vandorpe, & Beckers, 2007). Because of that, in this case we analyzed their predictive-value judgments.

² In our simulations, occasionally the hit or the false-alarm rates had values of 0 or 1. The z function for these values yields $-\infty$ and ∞ , respectively. To avoid this problem, we followed the correction suggested by Snodgrass and Corwin (1988). Specifically, in the computation of the hit rate, we added 0.001 to the number of hits and 0.002 to the number of outcome-present trials. The same correction was used in the computation of the false-alarm rate. This correction only makes a minimal difference in the value of d' , except when either the hit or the false-alarm rates have extreme (0 or 1) values.

³ All the effect sizes included in the meta-analyses were computed from the t -values reported in the studies following the equations suggested by Lakens (2013), even when this resulted in d values slightly different from those reported by the authors. The random-effects meta-analysis was conducted using the ‘metafor’ R package (Viechtbauer, 2010).

1A		allergy	~allergy	
medicine	<i>a</i>	<i>b</i>		$p(o c) = a / (a + b)$
~medicine	<i>c</i>	<i>d</i>		$p(o \sim c) = c / (c + d)$

1B	o	~o	
c	80	20	
~c	20	80	
	$\Delta p = .60$		

1C	o	~o	
c	20	80	
~c	80	20	
	$\Delta p = -.60$		

1D	o	~o	
c	50	50	
~c	50	50	
	$\Delta p = .00$		

1E	o	~o	
c	80	20	
~c	80	20	
	$\Delta p = .00$		

1F	o	~o	
c	50	50	
~c	5	5	
	$\Delta p = .00$		

1G	o	~o	
c	80	20	
~c	8	2	
	$\Delta p = .00$		

Figure 1. Panel 1A represents a standard 2 x 2 contingency table. Panels 1B-1G represent examples of contingency tables yielding different Δp values.

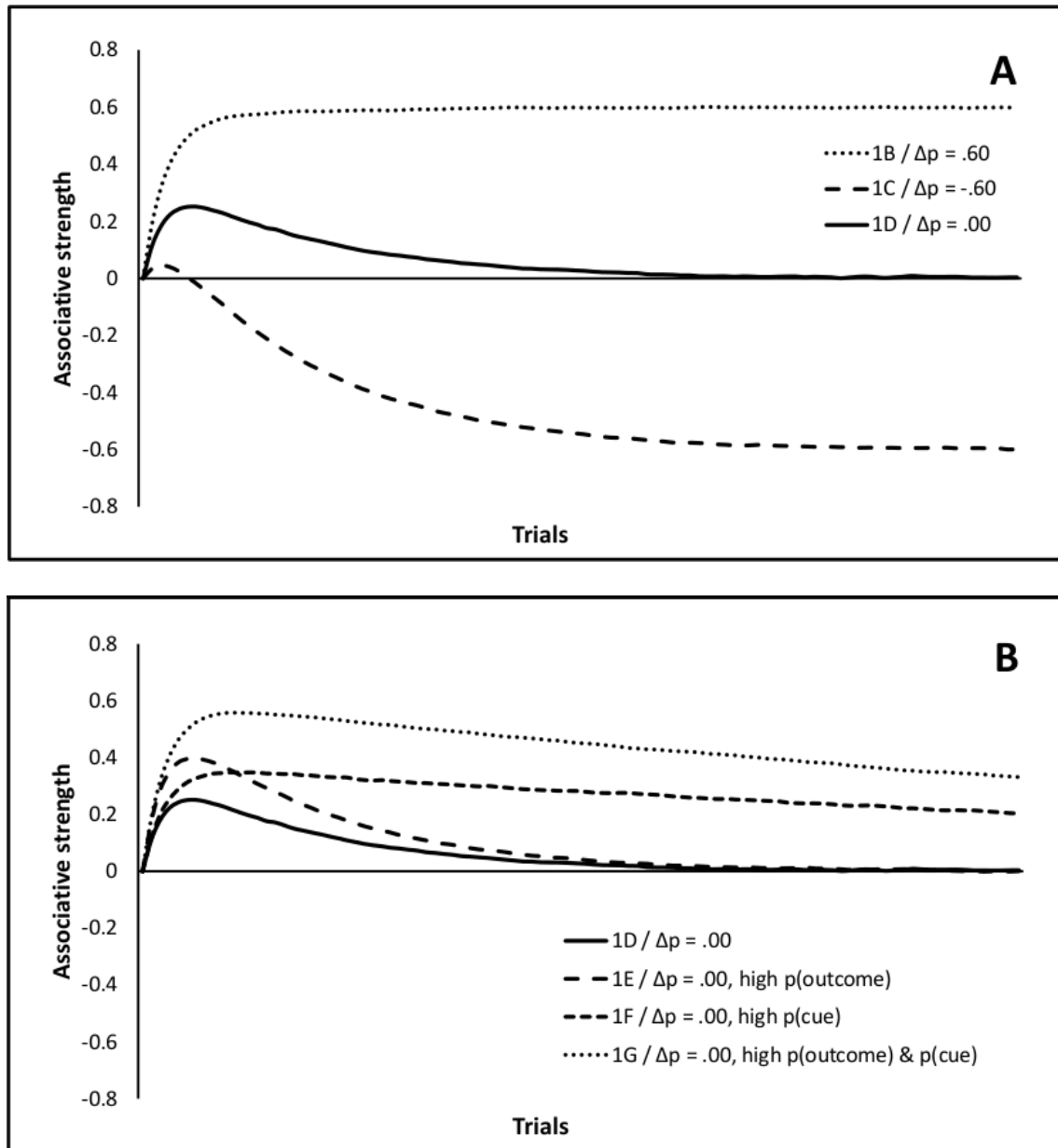


Figure 2. Results of a computer simulation of the six contingencies represented in Figures 1B-1G using the Rescorla-Wagner learning algorithm. The simulation was conducted using the Java simulator developed by Alonso, Mondragón, and Fernández (2012). For this simulation, the learning rate parameters were set to $\alpha_{\text{cue}} = 0.3$, $\alpha_{\text{context}} = 0.1$, $\beta_{\text{outcome}} = \beta_{\sim\text{outcome}} = 0.8$.

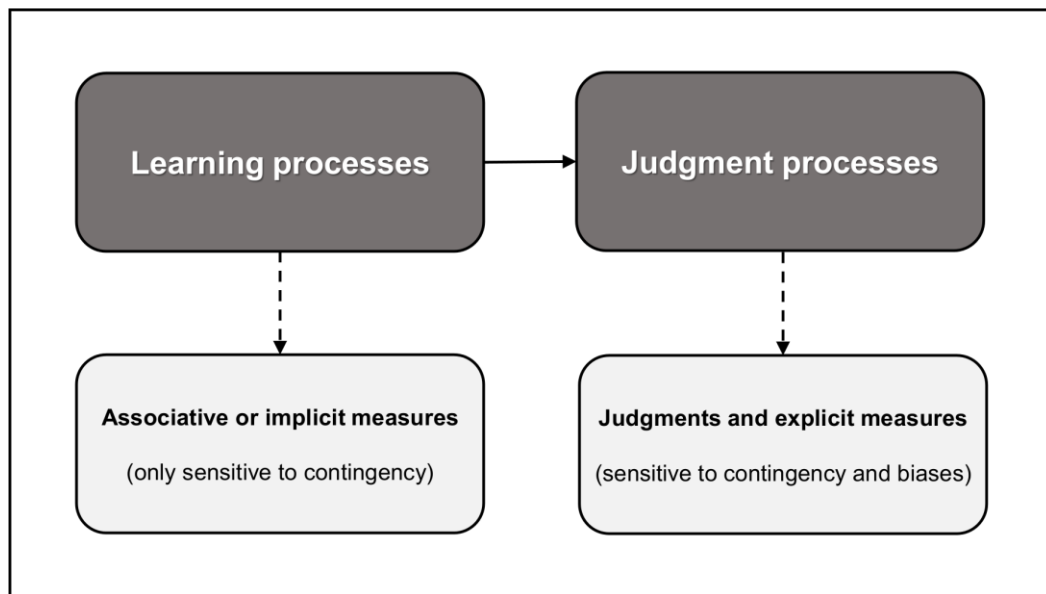


Figure 3. Schematic representation of dual-process models of biased contingency detection.

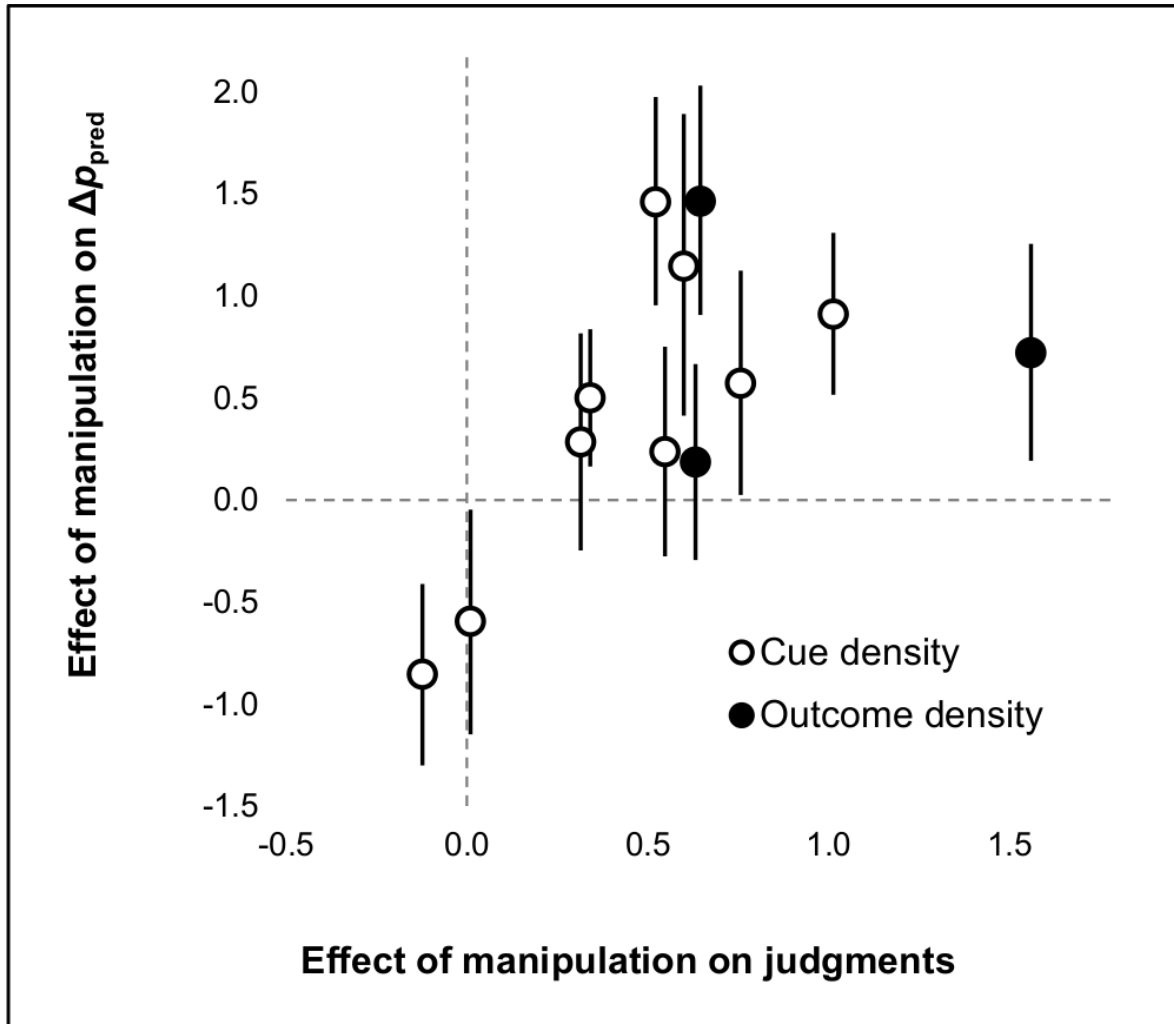


Figure 4. Scatterplot of effect sizes (Cohen's d) of the cue- and outcome-density manipulations on the Δp_{pred} index and on judgments in 12 experimental conditions. Error bars denote 95% confidence intervals.

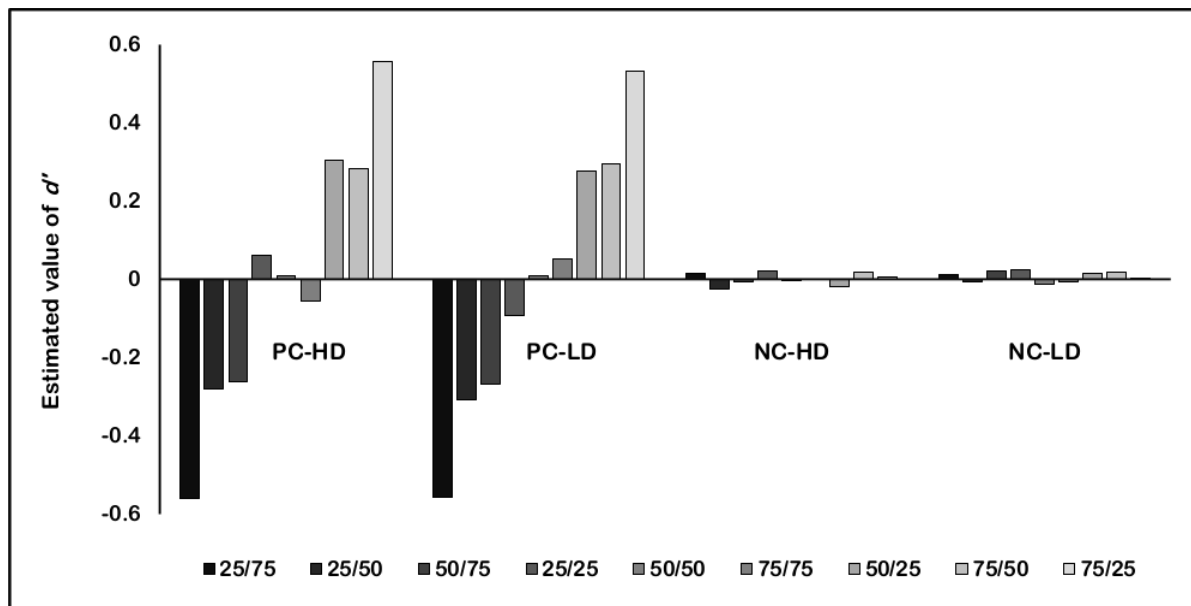


Figure 5. Simulated d' scores of participants with different response strategies in the four experimental conditions included in Perales et al. (2005, Experiment 1). Error bars denote the standard error of the mean across simulations.

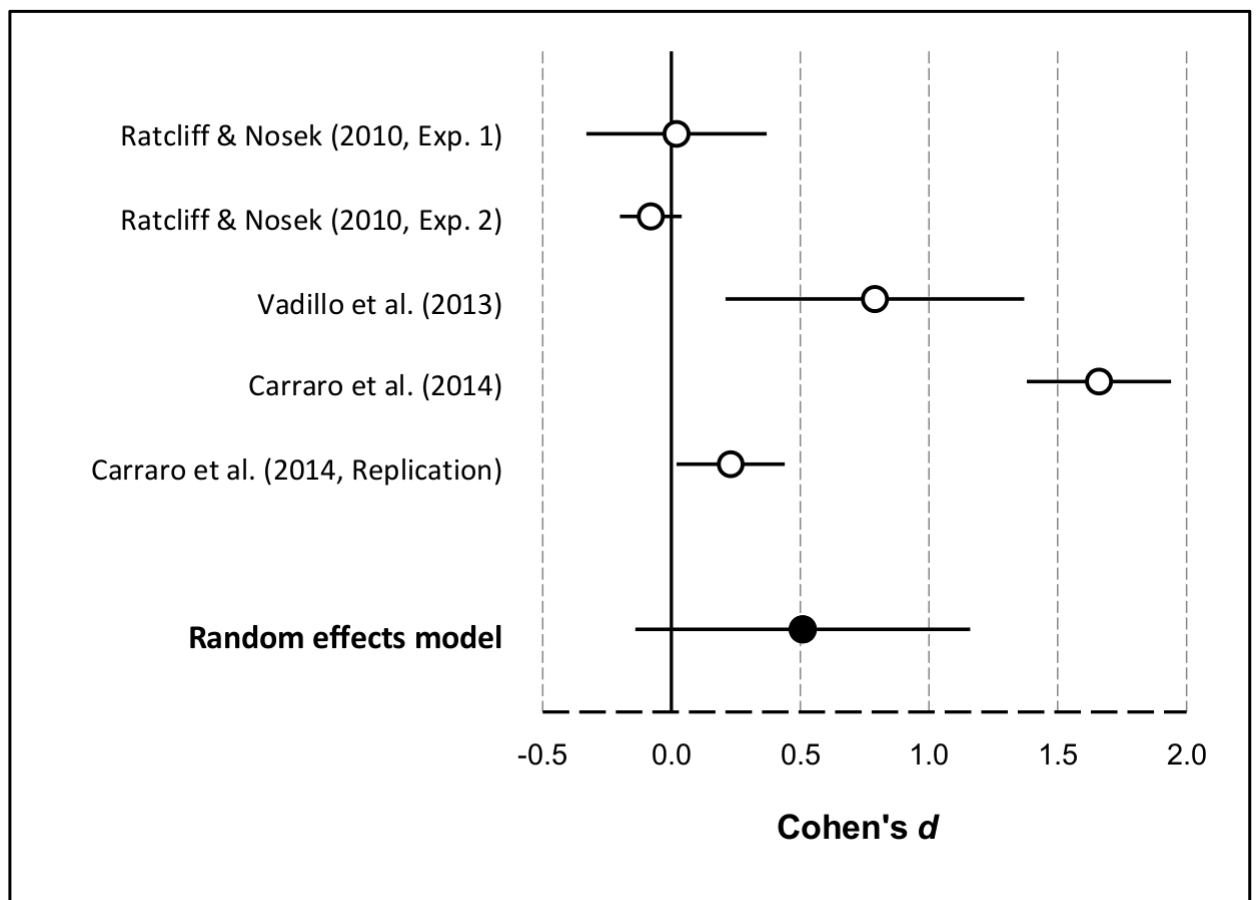


Figure 6. Forest plot of a meta-analysis exploring the results of the experiments that have measured illusory correlation effects in the IAT. Error bars denote 95% confidence intervals.